EE5907/EE5027 Week 2: Probability Review

The following questions are from Kevin Murphy's (KM) book "Machine Learning: A Probabilistic Perspective".

Exercise 2.6: Conditional independence

(a) Let $H \in \{1, \cdots, K\}$ be a discrete random variable, amd let e_1 and e_2 be the observed values of two other random variables E_1 and E_2. Suppose we wish to calculate the vector

$$\vec{P}(H|e_1, e_2) = (P(H = 1|e_1, e_2), \cdots, P(H = K|e_1, e_2)) \tag{1}$$

Which of the following sets of numbers are sufficient for the calculation?

i. $P(e_1, e_2), P(H), P(e_1|H), P(e_2|H)$

ii. $P(e_1, e_2), P(H), P(e_1, e_2|H)$

iii. $P(e_1|H), P(e_2|H), P(H)$

(b) Now suppose we now assume $E_1 \perp E_2|H$ (i.e., E_1 and E_2 are conditionally independent given H). Which of the above 3 sets are sufficent now?

Show your claculations as well as giving the final result. Hint: use Bayes rule.

Exercise 2.7: Pairwise independence does not imply mutual independence

We say that two random variables are pairwise independent if

$$p(X_2|X_1) = p(X_2) \tag{2}$$

and hence

$$p(X_2, X_1) = p(X_1)p(X_2|X_1) = p(X_1)p(X_2) \tag{3}$$

We say that n random variables are mutually independent if

$$p(X_i|X_S) = p(X_i) \ \forall S \subseteq \{1, \cdots, n\} \setminus \{i\} \tag{4}$$

and hence

$$p(X_{1:n}) = \prod_{i=1}^{n} p(X_i) \tag{5}$$

Show that pairwise independence between all pairs of variables does not necessarily imply mutual independence. It suffices to give a counter example.

Exercise 2.8: Conditional indepence iff joint factorizes

In the text we said $X \perp Y|Z$ iff

$$p(x, y|z) = p(x|z)p(y|z) \tag{6}$$

for all x, y, z such that $p(z) > 0$. Now prove the following alternative definition: $X \perp Y|Z$ iff there exist function g and h such that

$$p(x, y|z) = g(x, z)h(y, z) \tag{7}$$

for all x, y, z such that $p(z) > 0$

EE5907/EE5027 Week 2: MLE + MAP

The following questions are from Kevin Murphy's (KM) book "Machine Learning: A Probabilistic Perspective".

Exercise 3.1 MLE for the Bernoulli/ binomial model

Derive

$$\hat{\theta}_{MLE} = \frac{N1}{N} \tag{1}$$

by optimizing the log of the likelihood in Eq. (2)

$$p(\mathcal{D}|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{2}$$

Exercise 3.6 MLE for the Poisson distribution

The Poisson pmf is defined as $\text{Poi}(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$, for $x \in \{0, 1, 2, \cdots\}$ where $\lambda > 0$ is the rate parameter. Derive the MLE.

Exercise 3.7 Bayesian analysis of the Poisson distribution

In the previous exercise, we defined the Poisson distribution with rate λ and derived its MLE. Here we perform a conjugate Bayesian analysis.

a. Derive the posterior $p(\lambda|\mathcal{D})$ assuming a conjugate prior $p(\lambda) = Ga(\lambda|a, b) \propto \lambda^{a-1}e^{-\lambda b}$. Hint: the posterior is also a Gamma distribution.

b. What does the posterior mean tend to as $a \to 0$ and $b \to 0$? (Recall that the mean of a $\text{Ga}(a, b)$ distribution is a/b.)

Exercise 3.12 MAP estimation for the Bernouli with non-conjugate Priors

We discussed Bayesian inference of a Bernoulli rate parameter with the prior $p(\theta) = Beta(\theta|\alpha, \beta)$. We know that, with this prior, the MAP estimate is given by

$$\hat{\theta} = \frac{N_1 + \alpha - 1}{N + \alpha + \beta - 2} \tag{3}$$

where N_1 is the number of heads, N_0 is the number of tails, and $N = N_0 + N_1$ is the total number of trials.

Now consider the following prior, that believes the coin is fair, or is slightly biased towards tails:

$$p(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0.5 \\ 0.5 & \text{if } \theta = 0.4 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Derive the MAP estimate under the prior as a function of N_1 and N.